Climate Forecast Analysis Tools Framework

Núria Pérez-Zanón[#1], An-Chi Ho[#2], Francesco Benincasa[#3], Pierre-Antoine Bretonnière[#4], Louis-Philippe Caron[*5], Chihchung Chou[#6], Carlos Delgado-Torres[#7], Llorenç Lledó[#8], Nicolau Manubens[#9], Lluís Palma[#10]

[#]*Earth Science Department, Barcelona Supercomputing Center (BSC)*

[1]nuria.perez@bsc.es, [2]an.ho@bsc.es,
[3]francesco.benincasa@bsc.es,[4]pierre-antoine.bretonniere@bsc.es,[6]chihchung.chou@bsc.es,[7]carlos.delgado@bsc.es,[8]llledo@bsc.es,[9]nicolau.manubens@bsc.es,[10]lluis.palma@bsc.es,

[*]*Ouranos, 550 Sherbrooke St W, Montreal, Quebec H3A 1B9, Canada*

[5]caron.louis-philippe@ouranos.ca

***Keywords*— Climate Services, Climate forecast verification, Forecast calibration, Climate analysis tools**

Extended Abstract

The climate forecast analysis tools provide functions implementing the steps required for the analysis of sub-seasonal, seasonal and decadal forecast and operational climate services, allowing researchers to manipulate climate data and apply state-of-the-art methods taking advantage of the high-performance computational resources. Researchers can share their methods while reducing development and maintenance cost. An ecosystem of R packages covering these needs is under continuous development.

A. Introduction

Initialized Earth system predictions are made by starting a numerical prediction model in a state consistent with the observations corresponding to a specific day but adding perturbations on the initial conditions to generate an ensemble of simulations. The simulations run forward in time for sub-seasonal (around 30 days ahead), seasonal (from the 3 months to a year) and decadal (around ten years ahead). For each forecast horizon, the Earth system models are adapted to improve the representation of the environmental conditions that have a bigger impact on their predictability. Apart from the accuracy of initial atmospheric condition, the sources of predictability in sub-seasonal forecast comes from the stratospheric and surface representation whereas seasonal forecast is strongly affected by the sea surface temperature conditions (White et al., 2017). The source of predictability in decadal predictions is also affected by low-frequency modes on the extra-tropics sea surface temperature. However, the sub-seasonal to decadal communities share common scientific and technical challenges: initialization shock and drift; understanding the onset of model systematic errors; bias correction, calibration, and forecast quality assessment; model resolution; sources and expectations for predictability; and linking research, operational forecasting, and end-user needs (Merryfield et al., 2020).

At the Earth Sciences department of the Barcelona Supercomputing Center, the expertise in climate forecast research has traditionally been compiled in the s2dverification R package (Manubens et al., 2018) since its first release in 2009. The package provides tools implementing the steps required for the analysis of climate forecast, allowing researchers to share their methods while reducing development and maintenance cost. New packages have been developed to benefit from the available computational resources allowing researchers to conduct analysis that implies big data size and reducing the computation time.

Currently, 8 R packages (Table 1) are being maintained by the department and developed with contributions from external collaborators in the framework of European projects. In the next section, a description of the packages will be provided, whereas section 3 explains the flexibility and processing options. Finally, conclusions and future developments are summarized.

TABLE I
List and Descirption of Packages in the Climate Forecast Analysis Tools

Package name	Short description and link to CRAN
easyNCDF	Read/write netCDF files into/from multidimensional R array. https://CRAN.R-project.org/package=easyNCDF
multiApply	Apply functions to multiple multidimensional arrays or vectors allowing parallel computation https://CRAN.R-project.org/package=multiApply
s2dverification	Functions for Forecast Verification and visualization https://CRAN.R-project.org/package=s2dverification
s2dv	Adaptation of s2dverification to multiApply https://CRAN.R-project.org/package=s2dv
CSTools	Methods for forecast calibration, statistical and stochastic downscaling, optimal forecast combination and tools to obtain tailored products. https://CRAN.R-project.org/package=CSTools
CSIndicators	Sectorial Indicators for Climate Service (under-development in internal gitlab project)
ClimProjDiags	Climate extreme indices, evaluation of the agreement between models, weight and combination functions. https://CRAN.R-project.org/package=ClimProjDiags
startR	Data retrieval and processing tools https://CRAN.R-project.org/package=startR

B. Methods in the Climate Forecast Analysis Tools

The methods included in the climate forecast analysis tools aim to obtain a research result or climate service product from the climate forecast and reference datasets (e.g. reanalysis) by transforming the data and applying state-of-the-art methods:

- Different approaches to understand the links between climate variability and their impacts are explored by the researchers. That is the case, for instance, of the Madden-Julian Oscillation (MJO), a prominent feature of the tropical atmospheric circulation at sub-seasonal time scales, is known to modulate atmospheric variability in the Euro-Atlantic region (Lledó et al., 2020). Several atmospheric circulation patterns are already defined in the scientific literature and they are calculated by applying different methodologies (e.g.: Empirical Orthogonal Functions or area-weighted means).
- Skill metrics to assess the quality of their forecasts by comparing them against reference observation datasets can be deterministic, probabilistic and multivariate.
- Forecast post-processing refers to scientific methods to increase the quality of the forecasts, such as calibration methods to remove the systematic model error, or to increase the usability of the forecasts, such as downscaling techniques that allow achieving a finer resolution than the coarse original model resolution.

● Apart from essential climate variables, tailored indicators can help narrow the usability gap between pure science and stakeholders of key socio-economic sectors, such as renewable energy, air quality, agriculture or insurance, as well as to the general public.

● Visualization tools are a fundamental step to explore and communicate results in scientific writing and to end-users.

After reading the data from file to RAM, there is not a unique possible combination of methods, the users can skip some processes or apply multiple methods to the same data.

C. Tools Flexibility

Originally, s2dverification package relied on a fixed structure of the data: an array (i.e. multi-dimensional object) in which each of the dimension corresponds to dataset identifier, member of the ensemble, start date for the initialization forecast date, lead time for the forecast time step, latitude and longitude positions defined in the data retrieval step. This structure is suitable for most of the transformations and methods, but several other needs may require extra dimensionalities, such as atmospheric level or weather type identification.

In this context, multiApply package was released as a variant of apply functions extending this paradigm. Its only function, Apply, efficiently applies functions taking one or a list of multiple unidimensional or multidimensional arrays (or combinations thereof) as input. This saves development time by preventing the R user from writing often error-prone and memory-inefficient loops dealing with multiple complex arrays. Also, a remarkable feature of Apply is the transparent use of multi-core through its parameter 'ncores'. Therefore, the latest packages (i.e. s2dv, CSTools and CSIndicators) are developed taking advantage of multiApply, and guidelines on how to develop functionalities have been written to be followed autonomous by researchers.

For the data retrieval and datasets processing steps, startR package has been developed to provide a flexible way to retrieve data from files, as well as, performing analysis when the size of involved data overcomes the available RAM memory by implementing the MapReduce paradigm, chunking the data and processing them either locally or remotely on high-performance computing systems, leveraging multi-node and multi-core parallelism where possible (Pérez-Zanón et al., 2021).

The result is that functions in the climate forecast analysis tools framework use multi-dimensional arrays with named dimensions, the users can set up parameters specifying the dimension(s) in which the functions should be applied, as well as, the number of cores to use in the computation. The learning curve is expected to show a big positive trend once the user gets an understanding of the multi-dimensional array with named dimensions as the main object in which data is stored.

D. Conclusion and Future Enhancement

The success of the climate forecast analysis tools is proved by the number of peer-reviewed articles published by the department researchers, operational climate services deployed and projects successfully undertook. The tools are under continuous development exploring new methods and allowing extra transformations in a user-friendly way. Users support, internal training and discussions, as well as, dissemination are a priority at this stage.

E. ACKNOWLEDGEMENTS

Part of this work was performed in the framework of the MEDSCOPE (MEDiterranean Services Chain based On climate PrEdictions) ERA4CS project (grant agreement no. 690462) funded by the European Union, the Copernicus Climate Change Service (C3S) (Metrics and Access to Global Indices for Climate Projections, MAGIC) and MEDGOLD (Turning climate-related information into added value for traditional MEDiterranean Grape, OLive and Durum wheat food systems, agreement no. 776467). We acknowledge all contributors to these tools.

References

[1] L. Lledó, F.J. Doblas-Reyes, "Predicting daily mean wind speed in Europe weeks ahead from MJO status" Mon. Weather Rev. ,2020. DOI: 10.1175/mwr-d-19-0328.1.

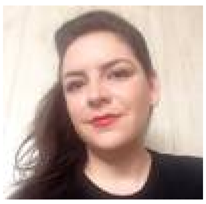
[2] N. Manubens, L-P. Caron, A. Hunter, O. Bellprat, E. Exarchou, NS. Fu Ckar, J. Garcia-Serrano, F. Massonnet, MM. En Egoz, V- Sicardi, E.C. Batt , C. Prodhomme, V. Torralba, N. Cortesi, O. Mula-Valls, K. Serradell, V. Guemas, F.J. Doblas-Reyes. "An R package for climate forecast verification" Environ. Model. Softw., 103: 29–42, 2018 DOI: 10.1016/j.envsoft.2018.01.018.

[3] W.J. Merryfield, J. Baehr, L. Batté, E.J. Becker, A.H. Butler, C.A.S. Coelho, G. Danabasoglu, P.A. Dirmeyer, F.J. Doblas-Reyes. "Current and emerging developments in subseasonal to decadal prediction". Bulletin of the American Meteorological Society. American Meteorological Society 101(6): E869–E896. 2020 DOI: 10.1175/BAMS-D-19-0037.1.

[4] N. Pérez-Zanón, A. Ho, N. Manubens, F. Benincasa, P. Bretonnière. "startR: A tool for large multi-dimensional data processing" the 8th BSC Doctoral Symposium 2021(submitted)

[5] C.J. White, H. Carlsen, A.W. Robertson, Klein RJT, Lazo JK, Kumar A, Vitart F. "Potential applications of subseasonal-to-seasonal (S2S) predictions". Meteorol. Appl., 315–325. 2017 DOI: 10.1002/met.1654.

Author biography



Núria Pérez-Zanón is a postdoctoral researcher and member of the Data and Diagnostic teams in the Computational Earth Sciences group at BSC. With a background in Physics and Meteorology (degrees from the University of Barcelona; UB), she obtained her PhD from Rovira i Virgili University (URV) in 2017 on Climate variability and Change detection in the central Pyrenees using instrumental and paleoclimate proxy data. During the thesis, one year and a half of research stay in LOCEAN (Laboratoire d'Océanographie et du Climat). She is the author of 14 peer-reviewed articles in international journals. She is currently coordinating the development and maintenance of the R tools of the department.